JOURNAL OF GEOPHYSICAL RESEARCH, VOL. 104, NO. D22, PAGES 27,545–27,564, NOVEMBER 27, 1999

Choosing meteorological input for the global modeling initiative assessment of high-speed aircraft

A. R. Douglass,[1] M. J. Prather,[2] T. M. Hall,[3] S. E. Strahan,[4] P. J. Rasch,[5] L. C. Sparling,[6] L. Coy,[4] and J. M. Rodriguez[7]

Abstract. The global modeling initiative (GMI) science team is developing a three-dimensional chemistry and transport model (CTM) for use in assessment of the atmospheric effects of aviation. This model must be documented, be validated against observations, use a realistic atmospheric circulation, and contain numerical transport and photochemical modules representing atmospheric processes. The model must retain computational efficiency for multiple scenarios and sensitivity studies. To meet these requirements, a facility model concept was developed in which the different components of the CTM are evaluated separately. The assessment of the impact on the stratosphere of the exhaust of supersonic aircraft will depend strongly on the meteorological fields used by the CTM. Three data sets for the stratosphere were considered: the National Center for Atmospheric Research Community Climate Model (CCM2), the Goddard Earth Observing System data assimilation system, and the Goddard Institute for Space Studies general circulation model. Objective criteria were developed to identify the data set that provides the best representation of the stratosphere. Simulations of gases with simple chemical control were chosen to test various aspects of model transport. The data sets were evaluated and graded on their performance on these tests. The CCM2 meteorological data set has the highest score and was selected for GMI. This objective model evaluation establishes a physical basis for interpretation of differences between models and observations. Further, the method provides a quantitative basis for defining model errors, for discriminating between different models, and for ready reevaluation of improved models. This will lead to higher confidence in assessment calculations.

1. Introduction

The Atmospheric Effects of Stratospheric Aircraft (AESA) component of the National Aeronautics and Space Administration (NASA) High Speed Research Program (HSRP) seeks to assess the impact of a fleet of high-speed civil transport (HSCT) aircraft on the lower stratosphere. There are several components to such an assessment. Measurements in the field and laboratory, characterization of the exhaust products, and development of realistic scenarios for the distribution of emissions all play important roles. Ultimately, the use of models to calculate the fate of aircraft exhaust, the buildup of such pollution in the lower stratosphere, and the model response of ozone to the change in lower stratospheric composition is a key element of the assessment, because models are the primary tools through which the impact on the ozone layer is quantified.

Assessment calculations for stratospheric aircraft have centered on results obtained with two-dimensional (latitude altitude) models [*Prather et al.,* 1992; *Stolarski et al.,* 1995]. However, as has been pointed out from the start of AESA [*Douglass et al.,* 1991], there are aspects to this assessment which are more appropriately modeled in three dimensions. The aircraft are proposed to fly mainly in the Northern Hemisphere and always over the oceans with a high concentration of flight paths in identifiable oceanic corridors. Thus the pollutant source is zonally asymmetric and concentrated. The meteorology of the Northern Hemisphere stratosphere is

[1]NASA Goddard Space Flight Center, Greenbelt, Maryland.

[2]Department of Earth System Science, University of California, Irvine, California.

[3]Goddard Institute for Space Studies, New York.

[4]General Sciences Corporation, Laurel, Maryland.

[5]National Center for Atmospheric Research, Boulder, Colorado.

[6]Joint Center for Environmental Technology, University of Maryland, Baltimore County, Maryland.

[7]Rosenstiel School of Marine and Atmospheric Sciences University of Miami, Miami, Florida.

Paper number 1999JD900827.
0148-0227/99/1999JD900827$09.00

influenced by the land-ocean pattern, thus the transport of polluted air from the stratosphere to the troposphere is also asymmetric. There have been efforts to evaluate the importance of these asymmetries to the assessment calculation and to quantify expected differences from a two-dimensional calculation [*Douglass et al.,* 1993; *Rasch et al.,* 1994; *Weaver et al.,* 1995, 1996]. Although studies so far suggest fairly small impacts to the buildup of exhaust for three-dimensional (3-D) versus two-dimensional (2-D) models, uncertainties remained. There were differences in the pollutant buildup and the ozone response calculated by the various 2-D models, which focused attention on uncertainty in the 2-D model transport. The National Research Council panel on the AESA reviewed the NASA interim assessment [*Albritton et al.,* 1993], recognized this uncertainty, and recommended greater use of three-dimensional models, at least to evaluate the uncertainties associated with transport [*Graedel et al.,* 1994].

There are fundamental advantages to a three-dimensional calculation using state of the art representations of stratospheric chemical and transport processes, all of which could serve to reduce uncertainty in the assessment calculation, a primary goal of the AESA. Some of the uncertainties in the 2-D model assessments arise because there are observed aspects of stratospheric transport which are not represented explicitly in 2-D models. These include (but are not limited to) the wave mean flow interaction, the seasonal variation in the tropopause height, cross tropopause transport, and the seasonal evolution of the polar vortices. The 2-D models adopt diffusion coefficients to simulate the 3-D mixing of trace gases by planetary scale waves and to define the tropopause. While there is a strong theoretical basis for collapsing stratospheric transport to two dimensions, formulation of a 2-D model requires simplifying assumptions, and the formulations in current use are not unique, as evidenced by the range of results for model transport tests as reported by the Second Stratospheric Models and Measurements Workshop [*Park et al.,* 1999].

The 3-D models address these uncertainties by improving the physical basis for representing these processes. In some cases, comparisons of models with observations reflect these improvements. For example, the amplitude of the annual cycle in total ozone at northern middle latitudes is generally closer to the observed amplitude in 3-D models than in 2-D models [*Rasch et al.,* 1995; *Douglass et al.,* 1996]. The improved agreement is at least partially a result of a more physical representation of the tropopause and the concomitant transport in the lowermost stratosphere. It is important to remember that 2-D models have long been used to calculate constituent evolution, and comparison of calculated fields with zonal means of global observations has been a principal means of evaluating the 2-D models [e.g., *Prather and Remsberg,* 1993]. As noted above, model transport has a strong theoretical basis but retains a strong phenomenological component due to underlying simplifying assumptions and parameterizations. The 3-D models do not have this heritage for constituent modeling, and it is not likely that they will produce zonal mean results that compare better with zonal mean observations than 2-D models. However, improvement in the physical basis of the models sets the stage for physically based improvements in the 3-D model, often through interpretation of the differences between model fields and constituent observations.

The major disadvantage to utilizing 3-D models for assessment is their large computational requirement. Since the motivation for using the 3-D model rests on the improved physical basis of the model, the horizontal and vertical resolution must be adequate to resolve important transport processes. The transport and photochemical time steps must both be substantially smaller than the time steps often used in 2-D models. Model evaluation also becomes a larger task, requiring both computational and human resources. It is impractical that 3-D assessments follow the path of 2-D assessments, in which independent calculations are produced by several research groups. To gain the benefits of using the 3-D assessment and to maintain involvement of several research groups, the global modeling initiative (GMI) science team was formed.

The goal of this group is to produce a well-tested 3-D chemistry and transport model that is useful for assessment calculations. A modular design allows various numerical transport schemes, photochemical schemes, and sets of meteorological data (winds and temperatures) to be tested within a common framework. The choices of which scheme to use have been made considering both performance and computational requirements. The choice for the numerical scheme for photochemistry is given by P. S. Connell et al. (The global modeling initiative assessment model: Stratospheric photochemistry mechanism and solver, manuscript in preparation, 1999); transport is discussed by D. A. Rotman et al. (The global modeling initiative assessment model: Model integration and testing of the transport shell, manuscript in preparation, 1999); the initial results from an assessment calculation are presented by D. E. Kinnison et al. (The global modeling initiative assessment model: Application to high speed civil transport perturbation, manuscript in preparation, 1999). This paper centers on establishing criteria to choose the meteorological input fields for the assessment model.

This exercise of evaluating different meteorological fields for use in the model sets an important precedent for modeling work. The models are evaluated strictly by comparisons with observations and are scored quantitatively relative to each other and to a standard defined by the observations. The approach given here provides a quantitative, reproducible method for evaluating a meteorological data set. Because the tests are themselves physically based, this provides groundwork for identification of needed model improvements. Application of

the same tests to improved models will give a quantitative measure of the improvements. In the future the tests may be applied to other meteorological data sets; the tests themselves may be refined, expanded, replaced, or augmented. However, the methodology should stand and serves as a challenge to move from subjective to quantitative model evaluation. An obvious extension to this application will be evaluation of the tropospheric and lower stratospheric transport produced by various meteorological data sets as part of the Atmospheric Effects of Aircraft Project (AEAP) subsonic assessment (SASS), also an objective of GMI.

The three candidate meteorological data sets are described in section 2. To distinguish among them, aspects of transport thought to be important to the stratospheric assessment were identified, and tracer simulations to examine these transport issues were designed. These simulations are described in section 3. The tests themselves and the model performance on the tests are presented in section 4. Some discussion and conclusions, including the choice of the meteorological data to be used in the assessment, are given in section 5.

2. Meteorological Input Fields

2.1. GMI-CCM2

The GMI-CCM2 meteorological fields (hereinafter referred to as CCM2) come from a middle-atmosphere version of the community climate model version 2 (MACCM2). The standard configuration of the community climate model [*Hack et al.*, 1994] is modified to produce a simulation appropriate for the stratosphere by reducing the horizontal resolution and increasing the vertical resolution. The middle atmospheric model is run at a horizontal resolution of about 2.8° latitude by 5.6° longitude, with 44 levels in the vertical extending from the surface to about 75 km. Near the tropopause the spacing between levels is about 1 km; the maximum vertical spacing is about 2.5 km. This model configuration provides a tropospheric simulation, similar to that produced using the the standard version of the model, and also a realistic stratospheric simulation. The meteorological fields compare well with observations of winds and temperatures in the Northern Hemisphere. However, the simulation of the Southern Hemisphere winter stratosphere does not compare well with observations. The simulated winter vortex is much stronger than observed, and the simulated polar temperatures are colder than observed from the upper troposphere to 1 hPa. These similarities and differences are discussed in detail by *Boville* [1995].

These meteorological fields were used in an off-line full chemistry simulation, and comparisons of the resulting constituent fields with observations are discussed by *Rasch et al.* [1995]. There are points of similarity and differences in comparisons of model values from this simulation with observations. For example, the zonal mean of model total ozone captures many of the features observed by the Total Ozone Mapping Spectrometer (TOMS). However, in the comparisons presented by *Rasch et al.* [1995] the horizontal gradients for CH_4 are generally steeper than observed; for example, at 10 hPa the September difference in CH_4 between the tropics and 45° latitude in the Northern (Southern) Hemisphere is 0.65 ppmv (0.8 ppmv) compared with 0.4 ppmv (0.45 ppmv) as observed by the Halogen Occultation Experiment (HALOE) on the Upper Atmosphere Research Satellite (UARS) [*Russell et al.*, 1993].

The standard CCM2 gravity wave drag parameterization, which strongly influences the model middle atmosphere circulation, assumed a zero phase speed gravity wave source originating solely from flow over sub-grid-scale orography [*Boville*, 1995]. This source of gravity waves was augmented to provide for a more general source of gravity waves with nonzero phase speed arising from a variety of sources (shear instabilities, frontal propagation and convection) by adding separate zonally uniform sources for Northern Hemisphere, Southern Hemisphere and equatorial gravity waves. As a result of this change, the errors associated with an excessively strong jet and cold temperatures in the Southern Hemisphere winter night are dramatically reduced. The resulting meteorological fields have been used for studies of the dynamics and photochemistry of the Southern Hemisphere [e.g. *Brasseur et al.*, 1997]. However, in the Northern Hemisphere winter, the revised version compares less well with meteorological data in that the temperature is somewhat too warm, stratospheric sudden warmings are produced too frequently, and the final warming occurs too early in the model spring.

Data sets from both configurations of the model were provided to the AESA effort. Because of the improvements to the Southern Hemisphere climatology which must affect the overall model circulation and transport, the meteorological winds from the revised version of MACCM2 were judged to provide a better representation of the stratosphere than the first version, in spite of the known biases between model and observations for the Northern Hemisphere winter. The tests described below utilize the data set with the revised gravity wave parameterization.

2.2. GMI-GISS

The Goddard Institute for Space Studies (GISS) meteorological fields used in GMI come from a middle-atmosphere version of the GISS general circulation/climate model [*Hansen et al.*, 1983]. Improvements to physical parameterizations used in the standard GISS climate model are detailed by *Rind and Lerner* [1996], who refer to the physics used in this version as "model II′." Briefly, the middle-atmosphere model used here has higher vertical resolution (31 layers) than the standard model, an upper boundary at 80 km, and a gravity wave drag parameterization that greatly improves the simulation of stratospheric tem-

peratures and the strength of the jets [*Rind et al.*, 1998, and references therein]. The vertical grid for the general circulation model uses sigma coordinates from the surface to 150 hPa and fixed pressure levels above to a model lid of 0.002 hPa. The layer thickness is about 2 km near 150 hPa and increases to about 4 km throughout the stratosphere. The horizontal resolution is 4° latitude by 5° longitude. The GMI-GISS (hereinafter referred to as GISS) chemical transport simulations use the 6 hour averaged wind and temperature fields recorded from a single year of this model. The top four layers (above 72 km) are combined into a single layer. Analysis of the dynamical properties of these wind fields is given by *Rind et al.* [1998] and references therein, and the use of these wind fields in a CTM is described by *Hannegan et al.* [1998].

2.3. GMI-GEOS DAS

The GEOS DAS (Goddard Earth Observing System data assimilation system) winds are produced by a stratospheric version of the GEOS-1 system described by *Schubert et al.* [1993]. The stratospheric version has the top raised to ~ 0.1 hPa with a total of 46 vertical levels. The horizontal resolution is 2° latitude by 2.5° longitude. The vertical winds are mapped to a 29 level grid using a scheme that minimizes errors due to noise in the assimilated wind field. This technique integrates horizontal divergence over a column and maps the horizontal winds onto new vertical levels, while conserving the mass flux (S. J. Lin, personal communication, 1998). Near the tropopause the vertical spacing is about 1 km; the spacing is about 2.5 km through the middle and upper stratosphere. The output fields are available 4 times per day.

Unlike the other data sets used in this study, the GMI-GEOS DAS winds and temperatures (hereinafter referred to as GEOS DAS) are based on observations; the period for the assimilation is October 1995 through September 1996. The observations include rawinsonde and satellite temperature retrievals, as well as surface observations and cloud track winds. A short GCM forecast (3 hours) provides the first guess for an optimal interpolation data analysis procedure. The period used here is part of a longer assimilation run begun in May 1995 for the Stratospheric Tracers of Atmospheric Transport aircraft mission.

Some aspects of stratospheric data assimilation are described by *Coy and Swinbank* [1997] and *Coy et al.* [1994]. The global coverage from the satellite observations produces a good representation of the temperature and mass field. The middle- and high-latitude winds associated with the vorticity field are well represented through the use of the geostrophic approximation. In the lower stratosphere, additional information is provided by the middle-latitude rawinsonde stations. At low latitudes the winds are less certain because there are not many rawinsondes that report tropical stratospheric data, and the uniformity of the tropical mass field makes the geostrophic approximation much less useful. The horizontal divergence of the wind field is not directly observed in the stratosphere. In the GEOS DAS products it is determined by the interaction of the general circulation model with the inserted observations. *Weaver et al.* [1993] showed that for a previous version of the assimilation system, the shock of data insertion led to an unrealistic residual circulation. Improvements to the assimilation system reduced this effect, and assimilated fields were shown to produce a credible annual cycle for total ozone [*Douglass et al.*, 1996]. *Coy and Swinbank* [1997] showed that for the assimilation fields used here, the GEOS DAS zonal mean divergence field produced a qualitatively correct residual mean meridional circulation; the tests described here evaluate the residual mean meridional circulation quantitatively.

3. Constituent Simulations

The primary goal of testing these wind fields is to evaluate their ability to produce realistic transport in the lower stratosphere. To this end, we performed several simulations of long-lived tracers and tested the resulting fields against observations. All meteorological fields were used on a 4° latitude by 5° longitude horizontal grid; for some this required interpolation. The number of vertical levels was maintained, unique to each model, as described in section 2. Numerical transport was accomplished using a scheme developed by *Lin and Rood* [1996]. Only advective transport was considered. This is acceptable for the stratosphere, but for tropospheric simulations, it will be necessary to include vertical transport by convection.

Nitrous oxide serves as a tracer of stratospheric transport. Its only sources are in the troposphere, and it is long lived in much of the stratosphere. Loss is prescribed from a photochemical model that used observed zonal mean climatologies for temperature and ozone. The loss frequency for N_2O (photolysis and reaction with $O(^1D)$) is tabulated as a zonal mean function of latitude, pressure, and month [*Hall and Prather*, 1995; *Avallone and Prather*, 1997]. Because N_2O is also the primary stratospheric source of reactive nitrogen NO_y, and because the observed correlation between NO_y and N_2O in the lower stratosphere is well defined and linear, the model NO_y behavior can be inferred from the model N_2O.

CO_2 serves as a tracer of upper tropospheric/lower stratospheric transport. It has no sinks in the stratosphere but a small source from the oxidation of CH_4. Air enters the stratosphere at the tropical tropopause, where the CO_2 mixing ratio has both an annual cycle (3-4 ppmv) and an annual trend (~ 1.5 ppmv). Thus the phase and attenuation of the CO_2 annual cycle in various regions of the lower stratosphere contains information about transport and mixing. The CO_2 simula-

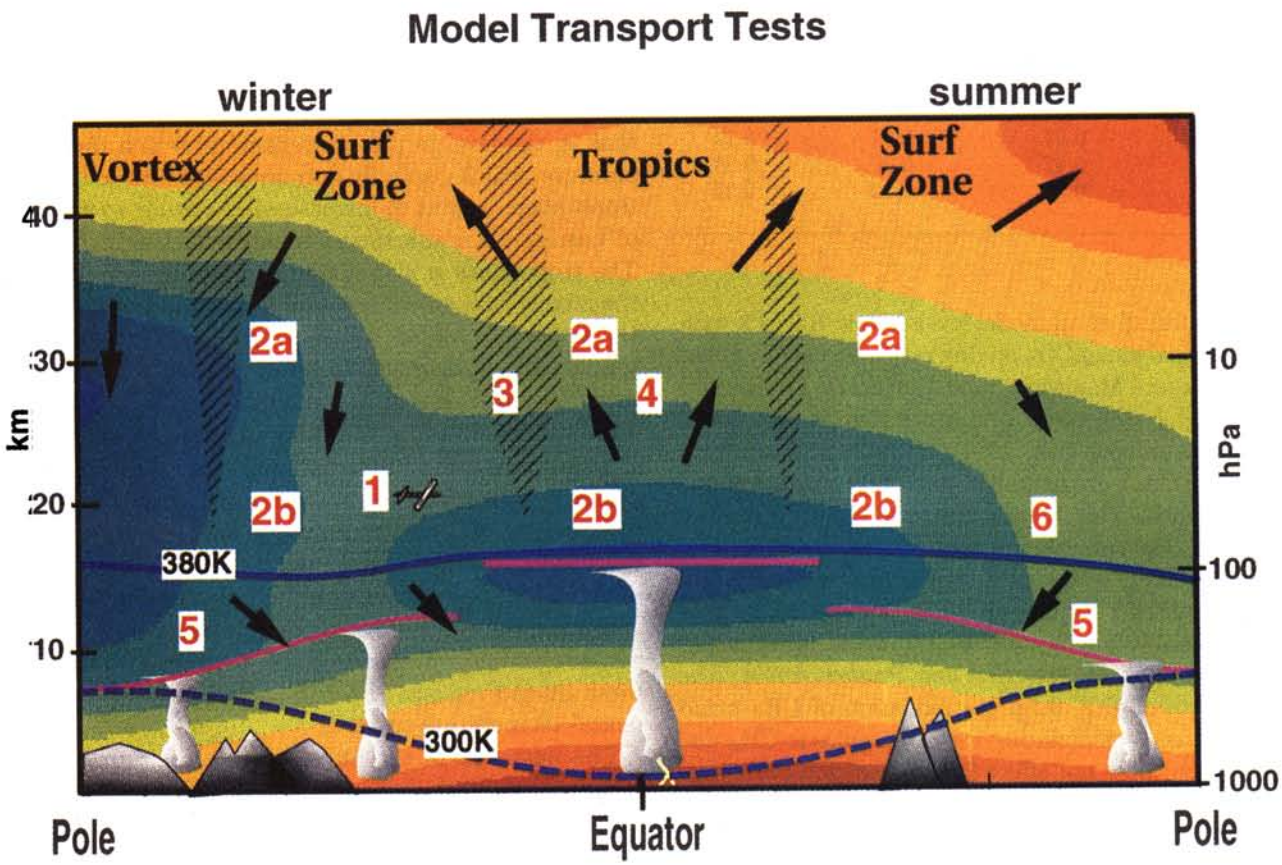

Plate 1. Schematic shows the principle features of stratospheric transport; the numerals indicate aspects of transport, which are the primary focus of a particular test. Test 1: the temperature distribution near the flight corridors; Tests 2a and 2b: the residual circulation and horizontal mixing; Test 3: separation of the tropics and middle latitudes; Test 4: large-scale ascent and horizontal mixing into the tropics; Test 5: separation of the troposphere from the lowermost stratosphere; Test 6: horizontal mixing from the tropics to the middle latitudes. The tests are described in detail in the text.

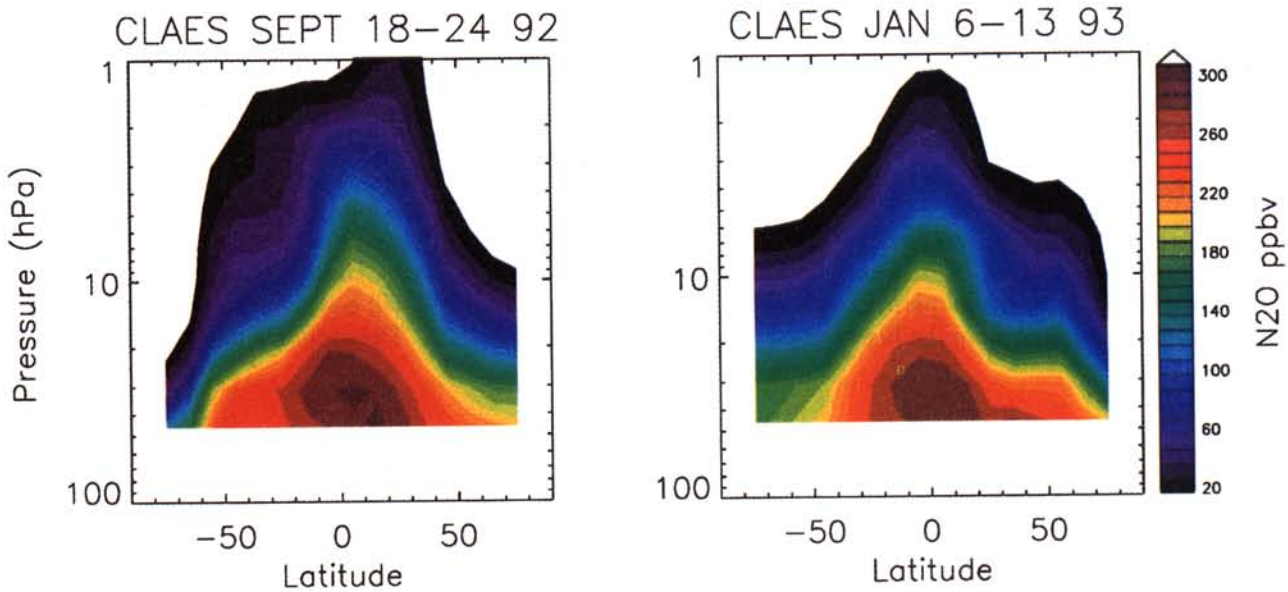

Plate 2. (left) Cryogenic Limb Array Etalon Spectrometer (CLAES) N_2O averaged in 10° latitude bins from September 18 to 24, 1992; (right) same as Plate 2 (left) but for January 6 to 13, 1993.

Table 1. Test 1 Grades (Temperature Test)

Model	40°N-50°N	60°N-70°N	Average
GEOS DAS	0.62	0.81	0.72
CCM2	0.47	0.26	0.36
GISS2	0.30	0.33	0.32

tions were forced at all model levels below 700 hPa with monthly zonal mean CO_2 mixing ratios derived from the NOAA Climate Monitoring and Diagnostic Laboratory (CMDL) global network of surface stations (P. Tans, personal communication, 1997). The CMDL data from remote stations were degraded to monthly values at 10° latitude bands. This 16 year record of CO_2 was used to force the CTM simulations by resetting the lower troposphere zonally every time step with the appropriate monthly value. This approach has been shown to produce reasonable simulations of CO_2 in the upper troposphere and stratosphere [*Hall and Prather,* 1993]. The small source of CO_2 from the oxidation of CH_4 could be calculated by simulating the stratospheric CH_4 distribution but does not affect the propagation of the tropospheric signal in CO_2 into the lower stratosphere and is neglected here.

Including a convective parameterization in the tropospheric constituent transport was found to produce too much mixing between the upper troposphere and the lowermost stratosphere for CCM2 and GEOS DAS (see test 5 below). Because stratospheric transport is of primary importance to the AESA assessment, the simulations considered here are run without using this parameterization; lack of convection is compensated for by applying the model lower boundary conditions up to the middle troposphere. Because subsonic aircraft fly near the tropopause in the upper troposphere and lower stratosphere, it will be important that the GMI model used in SASS include a realistic and validated parameterization of convective transport.

4. Tests

Plate 1 is a schematic representation of the stratosphere describing aspects of stratospheric transport which are important to model performance. The broad arrows on the diagram denote the Brewer Dobson circulation, which produces upward transport in the tropics and seasonally dependent downward transport in the middle and high latitudes. The thick solid line, with breaks in the subtropics, represents the tropopause; that is, the level in the atmosphere at which the temperature, which generally decreases with altitude in the troposphere, begins to increase with altitude as a result of heating from absorption of ultraviolet light by O_3. The tropopause is also marked by distinct changes in the mixing ratio of various constituents, including H_2O, O_3, and NO_y. A thick solid line indicates the 380 K potential temperature surface, which is approximately the lowest potential temperature surface that is fully in the stratosphere. Mixing can take place on constant potential temperature surfaces, but it is also inhibited by dynamical transport barriers. Shaded regions represent the subtropical boundaries and the boundaries between the middle latitudes and the winter polar vortex. As at the tropopause, constituent observations show distinct, seasonally dependent, horizontal gradients in these areas, suggesting separation of the air masses. Between these shaded regions, there is mixing on constant potential temperature surfaces.

The numerals in the plate indicate the part of the stratosphere to which a particular test is most relevant. The largest pollutant mixing ratios are expected at the flight level of the aircraft, marked by an airplane in the plate. Test 1 considers the temperature where the largest perturbation is expected. Test 2 has two parts, which examine the balance between the residual circulation and the horizontal mixing in the middle to upper stratosphere (Test 2a) and in the lower stratosphere (Test 2b). Test 3 concerns the separation between the tropics and the middle latitudes. Test 4 focuses on the vertical transport in the tropics and the effects of entrainment of air from the middle latitudes. Test 5 looks at transport near the tropopause. Test 6 concerns exchange between the middle latitudes and the tropics.

These tests are comprehensive but not exhaustive. Furthermore, in this first incarnation of the GMI model, we have chosen to include one "typical" year for each of the meteorological data sets. There is no effort to assess the role of the quasi-biennial oscillation or other phenomena contributing to longer timescale variability. We chose not to assess transport and mixing within the polar vortices; this will be considered further in the discussion of Test 2a. The accuracy of the model response

Table 2. Test 2a Grades (Middle to Upper Stratosphere Residual Circulation and Mixing)

Model	Annual Average	Horizontal Gradient	Vertical Gradient	Tropical Annual Cycle	Northern Annual Cycle	Southern Annual Cycle	Average
CCM2	0.61	0.33	0.83	0.13	0.14	0.17	0.37
GEOS DAS	0.22	0	0.42	0.06	0.04	0.10	0.14
GISS	0.44	0.08	0.58	0.10	0.08	0.06	0.22

Table 3. Test 4 Grades (Propagation of Annual Cycle)

Model	H_a (km)	c (mm s^{-1})	R	$g(R)$	$g(c)$	Mean Grade
CCM2	4.61	0.48	0.31	0.5	0.5	0.5
GEOS DAS	9.02	0.95	0.29	0.5	0	0.25
GISS	2.31	0.40	0.18	0	0.5	0.25
CO_2 H_2O	3.0	0.2	0.5			
HALOE (Mote)	7.4	0.34	0.7			
HALOE (Randel)	10.4	0.27	1.2			

to supersonic aviation (i.e., the HSCT assessment) is not likely to be equally sensitive to the model ability to match observations for each of these tests. However, the model performance on each of these tests provides an overall evaluation of the model transport, with the possibility that differences between the model and observations can be understood in terms of physical processes. The tests and model performance are discussed in detail below. Scores for Tests 1, 2a, 4, and 6 are given in Tables 1-4; the results for the ensemble of tests are given in Table 5.

4.1. Test 1 Temperature

The projected supersonic aircraft are expected to fly near 50 hPa, mostly at northern middle latitudes [*Prather et al.*, 1992]. Although the temperature per se does not affect the dispersion and buildup of aircraft exhaust, the processes affecting the impact of aircraft exhaust are temperature dependent, through gas phase photochemical reactions and through heterogeneous reactions on the surfaces of aerosols and polar stratospheric clouds (PSCs). In addition to the temperature dependence of some of the reactions that take place on aerosols and PSC surfaces (e.g., $ClONO_2$ + H_2O), the formation threshold for PSCs is temperature dependent, and the threshold itself will increase if the aviation products enhance the natural background of H_2O and HNO_3. This test focuses on the model climate at 50 hPa for two latitude bands, 40°N-50°N and 60°N-70°N. An annual climatology was compiled using 18 years of gridded global stratospheric analyses from the National Centers for Environmental Prediction (NCEP), formerly the National Meteorological Center (NMC). Temperatures at each level were interpolated to an equal area grid, so that all following calculations would be appropriately weighted. The monthly mean for the NCEP temperature includes all the equal area grid points within the specified latitude band for all 18 years; the monthly mean for the models includes only a single model year. The standard deviation for each month was calculated using daily temperature fields and their difference from their monthly mean. In addition, for the 18 years of NCEP analyses, a second standard deviation was calculated using the 18 monthly means for each year and their difference from the average of the 18 years. The first standard deviation measures the overall spread of temperatures seen on a daily basis for each model as well as the NCEP analyses, while the second standard deviation measures the interannual variability of the monthly means as seen in the 18 years of the NCEP analyses. The standard deviation was taken as a convenient numeric measure of variability, even though the daily temperature and the 18 years of NCEP monthly means are a small sample size for resolving a Gaussian distribution.

The annual cycles of the monthly mean and daily standard deviation of NCEP temperatures for the two latitude bands are compared with the annual cycle and daily standard deviation for temperatures from the CCM2, GEOS DAS, and GISS in Figure 1. At 40°N-50°N the GEOS DAS temperatures fall within 2 times the monthly mean standard deviation of the NCEP climatology (Figure 1, bottom left, bottom curve), as expected since the same data used to produce GEOS DAS are also used in one year of the NCEP data set. CCM2 temperatures are within 2 times the standard deviation of NCEP most of the year, falling outside only during spring and fall. The GISS temperatures are warmer than the NCEP climatology for most of the year. The daily standard deviations of the model fields (Figure 1, bottom left) generally follow the standard deviation of NCEP, except during winter when CCM2 and GISS show less variance than observed and GEOS shows slightly more for 1995 compared with the 18 year climatology.

At 60°N-70°N both CCM2 and GISS mean temperatures are 3° to 4° warmer than the NCEP climatology for most of the year. This is a known deficiency of the version of CCM2 used for this test, as discussed in section 2.1. The GEOS DAS temperatures show bet-

Table 4. Test 6 Tropical Midlatitude Transport

Model	CO_2 Amplitude	CO_2 Phase	Grade
CCM2	1.0	1.0	1.0
GEOS DAS	0.5	1.0	0.75
GISS	ng	ng	ng

Table 5. Average Scores

Model	GEOS DAS	CCM2	GISS
Test 1 temperature	0.72	0.36	0.32
Test 2a residual circulation (upper)	0.14	0.37	0.22
Test 2b residual circulation (lower)	0.63	0.68	0.43
Test 3 tropical-midlatitude separation	0.38	1	0.63
Test 4 propogation of annual cycle	0.25	0.5	0.25
Test 5 stratosphere troposphere separation	1	1	1
Test 6 tropical-midlatitude transport	0.75	1	ng
Average	0.58	0.73	0.51

ter agreement with NCEP temperatures at this latitude band than in the 40°N-50°N latitude band, falling within twice the NCEP monthly standard deviation (Figure 1, bottom right, bottom curve). Both CCM2 and GISS exhibit much less winter daily variance than NCEP (or GEOS DAS).

The differences seen in this figure are scored quantitatively using the following standard:

$$\text{grade}_\theta = 1 - \frac{1}{12}\sum_{i=1}^{12} \frac{\mid T_{\theta,i}^{\text{MODEL}} - T_{\theta,i}^{\text{NCEP}} \mid}{3\sigma_{\theta,i}^{\text{NCEP}}} \qquad (1)$$

where T is the monthly mean temperature, σ is the monthly standard deviation, θ refers to the latitude band, and the subscript i refers to each of the 12 months. Thus the grade is 1 if the model monthly mean is equal to the 18 year NCEP mean and high when the

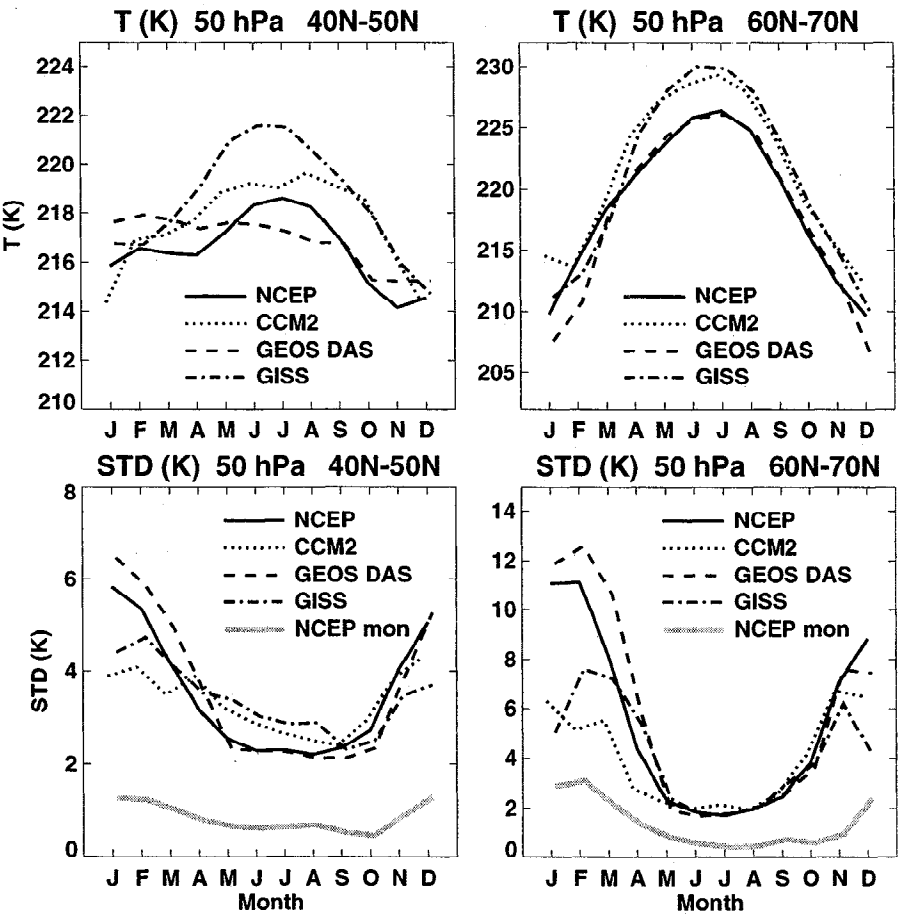

Figure 1. Annual cycles of the monthly mean and daily standard deviation of National Centers for Environmental Prediction (NCEP) temperatures for the two latitude bands are compared with the annual cycle and daily standard deviation for temperatures from CCM2, GEOS DAS, and GISS. Also shown are the standard deviation of the monthly averaged NCEP temperatures about the 18 year mean.

model difference from NCEP, weighted by the standard deviation of the monthly mean NCEP temperatures, is small. If the model difference from NCEP is greater than 3 times the standard deviation, the grade for that month is zero. Although it is important for strongly temperature dependent processes such as PSC formation that σ^{MODEL} is smaller than σ^{NCEP}, such processes cannot be realistic if there is a large difference from the mean temperature, and we have chosen not to score the comparison of the observed and modeled standard deviations. The scores for the three models are summarized in Table 1.

4.2. Test 2 Residual Circulation and Mixing

The strong vertical stability in the stratosphere constrains the vertical velocities to be a small fraction of the horizontal velocities. Thus while the horizontal winds in the lower stratosphere are of the order of 10 m s^{-1}, the average vertical velocities are measured as fractions of a millimeter per second ($\sim$ 0.001 m s^{-1}) [*Mote et al.*, 1998]. Since most trace gases in the stratosphere are vertically stratified, the vertical motions, though small, create horizontal gradients between regions of mean upward and downward motion. Concurrent horizontal motions tend to reduce these average horizontal gradients by mixing. Although at any given moment, large gradients may occur locally during mixing and wave breaking, for this evaluation we are concerned with monthly to seasonal time averages over 10° or greater latitudinal bands. For multiyear assessments the driving winds must realistically capture both the horizontal mixing and the slower vertical motions. In other words, since the CTM calculates vertical motion consistent with the horizontal wind divergence, the driving winds must capture both the divergence of the horizontal winds (for the slow upward motion) and the vorticity of the horizontal winds (for the horizontal mixing).

The above can be made more explicit by considering the residual mean meridional circulation and the Eliassen–Palm (EP) flux divergence associated with a given wind field [see *Andrews et al.*, 1987, p. 128]. An accurate representation of the EP flux divergence requires that the waves, seen mainly in the vorticity field, have realistic amplitudes and phase structures. The model waves must also dissipate by wave breaking or radiative damping at appropriate latitudes and heights. An accurate representation of the residual mean meridional circulation requires that the model or data assimilation system, which generates the winds, have an accurate radiative transfer scheme and a plausible subgrid-scale wave drag, in addition to an accurate representation of the EP flux divergence from the resolved waves. Realistic wave forcing is necessary for the downward control principle to produce a useful residual mean meridional circulation [*McIntyre*, 1992].

One way to judge the transport characteristics associated with a wind field is to study the long-lived tracer distribution produced by the wind field through advection and mixing. *Randel et al.* [1994], by evaluating the zonal mean tracer budget, show how the long-lived tracers depend on the residual mean circulation and eddy transport. These tests are designed to see how well the circulations associated with different driving wind fields can simulate the characteristics of an observed long-lived tracer.

4.3. Test 2a Middle to Upper Stratosphere

Nitrous oxide N_2O is emitted in the lower troposphere and destroyed in the stratosphere by photolysis and by reaction with excited atomic oxygen $O(^1D)$. The latter reaction is by far the smaller process through which N_2O is destroyed (about 10%) but is the dominant source of reactive nitrogen species in the stratosphere. The local chemical lifetime of N_2O decreases from hundreds of years at 100 hPa to several weeks in the upper stratosphere; the height and latitude dependence of the loss combined with the residual circulation and horizontal mixing produce the distribution seen in Plate 2. The N_2O pictured here was observed by the Cryogenic Limb Array Etalon Spectrometer (CLAES) on UARS [*Roche et al.*, 1993, 1996]. CLAES viewed from 34° latitude in one hemisphere to 80° latitude in the other hemisphere for periods of about 1 month during which the orbit precessed. The spacecraft was then yawed to keep the cold side away from the Sun, and CLAES viewed the opposite hemisphere. The data shown in Plate 2 were taken from several days before and after yaw maneuvers and thus provide near-global coverage. Plate 2 (left) shows data taken between September 18 and 24, 1992, and Plate 2 (right) shows data taken between January 6 and 13, 1993. Data are averaged in 10° latitude bands from 80°S to 80°N. These observations reveal seasonal differences in N_2O. For example, in the Northern Hemisphere the N_2O contours at middle latitudes are flatter during January than in September, and the high-latitude contours show evidence of winter descent.

CLAES operated for only 19 months, from October 21, 1991 to May 5, 1993. To minimize effects of Pinatubo aerosols, data for May 1992 to April 1993 are used for this evaluation. As noted by *O'Sullivan and Dunkerton* [1997], this winter corresponds to a westerly phase of the quasi-biennial oscillation (QBO) in the tropical lower stratosphere. The climatology of the general circulation models is more similar to the easterly phase of the QBO, and the assimilation year used by GMI is also an easterly year. There are systematic differences in the constituent distributions caused by differences in the residual circulation and in horizontal mixing [e.g., *O'Sullivan and Dunkerton*, 1997; *Randel et al.*, 1998; *Gray and Russell*, 1999]. The upwelling is stronger in the westerly phase, leading to higher mixing ratios above about 5 hPa in the tropics [*Randel et al.*, 1998]. There is significant movement of the lati-

tude of the maximum tracer gradient; in the Northern Hemisphere this is pushed equatorward at 14.7 hPa in the westerly phase relative to the easterly phase. We minimize the importance of this movement by comparing the model and observations in horizontal latitude bands which are 10° wide and centered on 45°N and 45°S (far from the subtropical edge) and the equator.

The September and January zonal mean distributions for the three models are compared with CLAES observations in successive panels of Plate 3. Inspection of the figures shows that all three models produce distributions that have many features in common with the CLAES observations. The maximum N_2O at any pressure is found in the tropics, and the N_2O decreases with altitude at all latitudes. The winter contours are flatter at middle latitudes in January than during September; their flat character indicates efficient planetary wave mixing. The upwelling region varies among the models and is narrowest for CCM2. There are also obvious differences among the models relative to each other and relative to CLAES data. These comparisons are quantified by requiring agreement of the model fields with several features in the CLAES data. The annual mean N_2O is calculated averaging all CLAES observations that fall within specified latitude and altitude (pressure) ranges. The vertical ranges are about 5 km, centered on 30, 10, and 2 hPa. For the annual cycle comparisons, all data within a month at the appropriate pressure and latitude band are averaged. Linear interpolation is used to fill observational gaps; at most, one month is missing.

HALOE observations of CH_4 are used to estimate systematic errors caused by comparing observations and model which are not in the same phase of the QBO. In the tropics, at 30 and 10 hPa, the average HALOE CH_4 mixing ratio varies by only a few percent. At 2 hPa the difference is larger but still less than 20%. At middle latitudes, there is no difference at 30 hPa. At 10 hPa and 2 hPa, mean differences are again less than 20%. The model value must be within 20% of the CLAES value for a score of 1 and within 40% of CLAES for a score of 0.5. The following comparisons are made: (1) annual mean at three levels for each latitude band (nine points of comparison), (2) the vertical gradients for each latitude band (six points of comparison), (3) the horizontal gradients for each pressure (six points of comparison), (4) the annual cycle in the tropics (36 points of comparison), (5) the annual cycle in the northern middle latitudes (36 points of comparison), and (6) the annual cycle in the southern middle latitudes (36 points of comparison). The scores on each component of the test are weighted to have a maximum of 1, and the overall score on this test is the average of the above six components. Results are given in Table 2.

A sense of these comparisons is gained by considering monthly average profiles for CLAES and for each model, shown in Plate 4 for the three latitude bands. Near the equator, GEOS DAS best represents the CLAES data in both magnitude and profile shape, and both GISS and CCM2 underestimate the CLAES N_2O profile. For both northern and southern middle latitudes, GEOS DAS profiles exceed observations; however, the profile shape is not consistent with observations. Between 40°N and 50°N the CCM2 and GISS profiles are nearly coincident with each other and with CLAES profiles. Between 40°S and 50°S, CCM2 and GISS profiles approximate the shape of the CLAES profiles, but the GISS profile is systematically high and CCM2 is systematically low. Given good comparisons for at least one latitude band for each of the meteorological fields, it is clear that there is no simple conceptual model of changes in the residual circulation or horizontal mixing that will resolve the discrepancies. It is also clear that the differences between the model and the observations cannot be attributed to QBO differences in the CLAES N_2O. For example, stronger upwelling in the model tropics would improve agreement of CCM2 and GISS with tropical CLAES profiles, but higher mixing ratios at middle latitudes would impact the good agreement there adversely. For all three models, the profile shapes are similar for the three latitude bands, whereas for CLAES, the tropical profile shape is distinctly different from the midlatitude profile shape. Note that HALOE CH_4 profiles also exhibit this difference in the shape of tropical and midlatitude profiles, regardless of the phase of the QBO. This qualitative discussion is quantified by the scores given in Table 2. CCM2 has the highest score overall. The relatively low scores are a good reflection of the general overall agreement suggested by Plate 3, but the important quantitative differences are shown in Plate 4.

4.4. Test 2b Lower Stratosphere

Strahan et al. [1999] have developed a lower stratospheric climatology for N_2O which is based on 175 flights of the ATLAS instrument on the NASA ER-2 between August 1988 and September 1997 [*Podolske and Loewenstein* 1993]. Plate 5 gives contour plots of this climatology in the northern fall and winter. The data shown here reflect many of the same features found in the CLAES data (Plate 2); N_2O is largest in the tropics and decreases with increasing potential temperature. The data reveal seasonal dependence; for example, lower mixing ratios of N_2O are seen in middle latitudes during the winter than in the fall.

For this application, mean profiles are calculated for three latitude ranges (35°S-50°S, 10°S-10°N; 35°N-55°N) and each of the four seasons, excepting 35°S-50°S during austral summer because there are no data for this time and location. These latitude ranges are chosen to approximate the middle-latitude surf zone, excluding both the polar vortex and the area of the subtropics where the horizontal gradients are steep. Because the polar vortex in the Southern Hemisphere extends more equatorward than in the Northern Hemisphere the latitudinal range for the Southern Hemisphere is 5° smaller

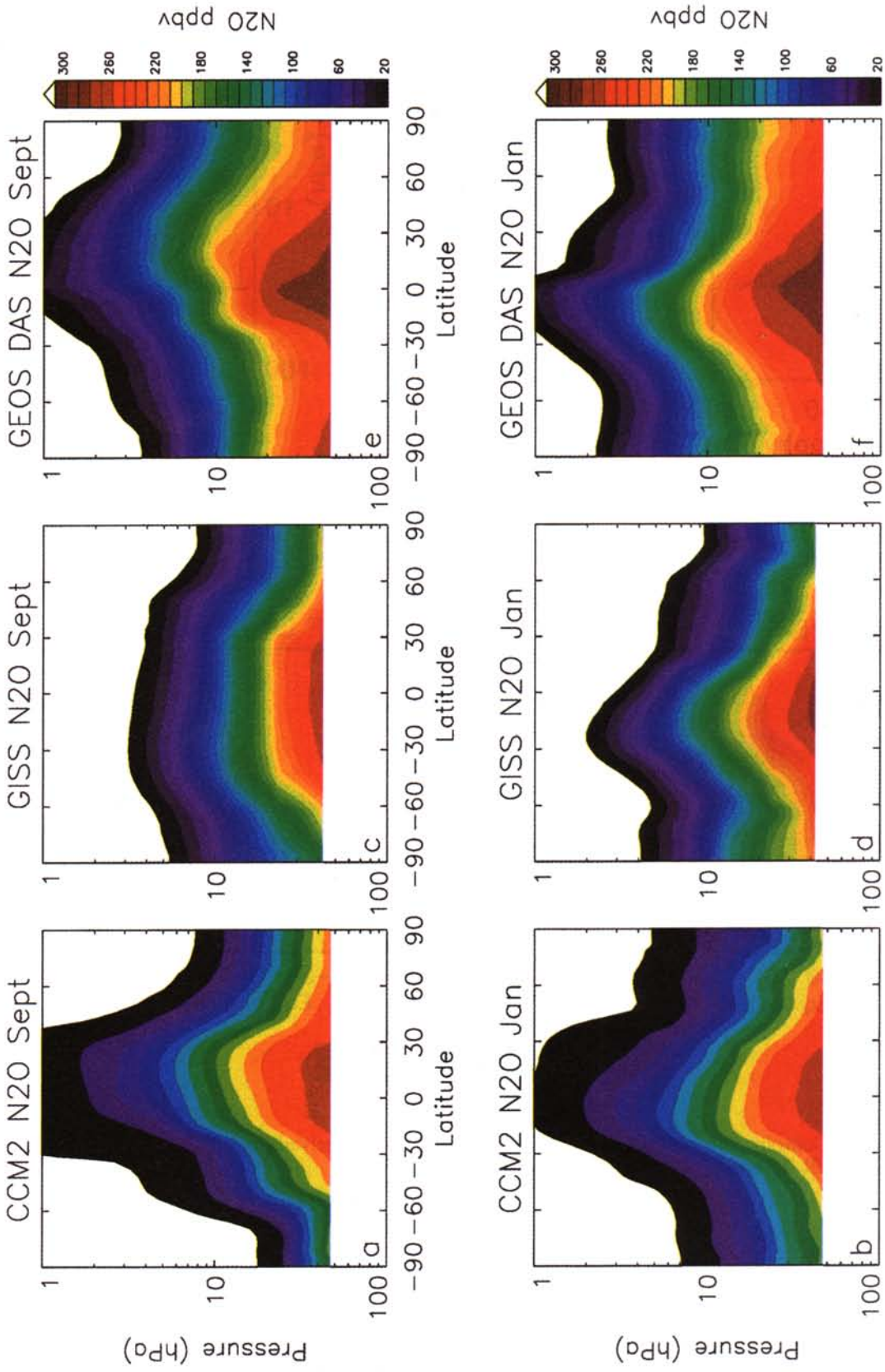

Plate 3. The September and January zonal mean distributions calculated using winds from GISS (a, b); for CCM2 (c, d); for GEOS DAS (e, f).

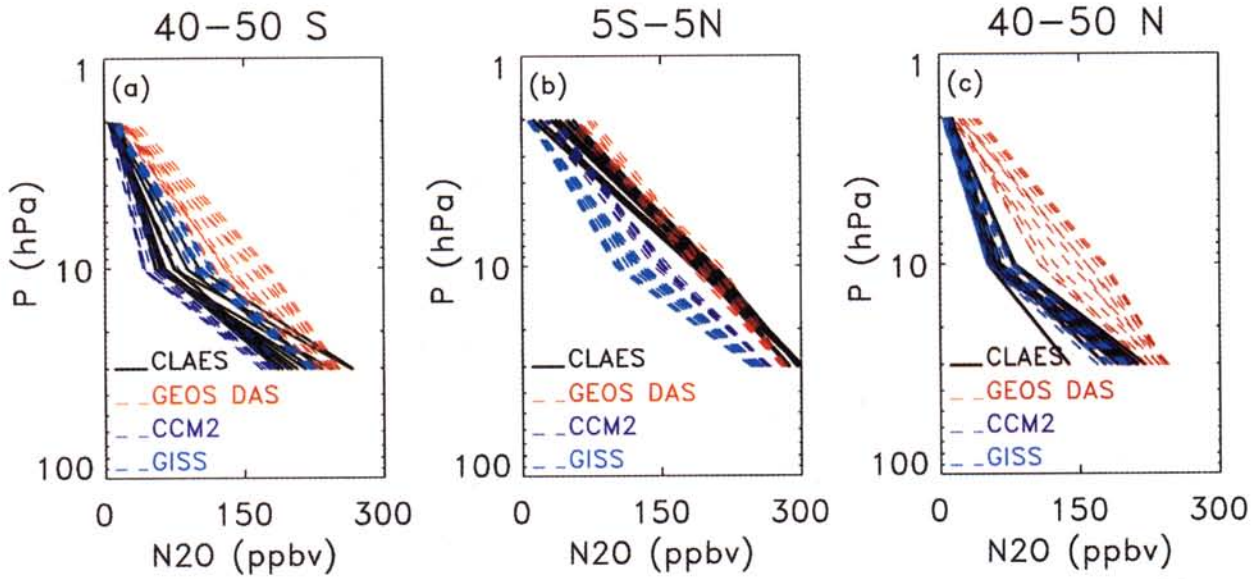

Plate 4. Mean N_2O profiles for CLAES and for each model for each month are shown for (a) 40°-50°S; (b) 5°S-5°N; (c) 40°-50°N.

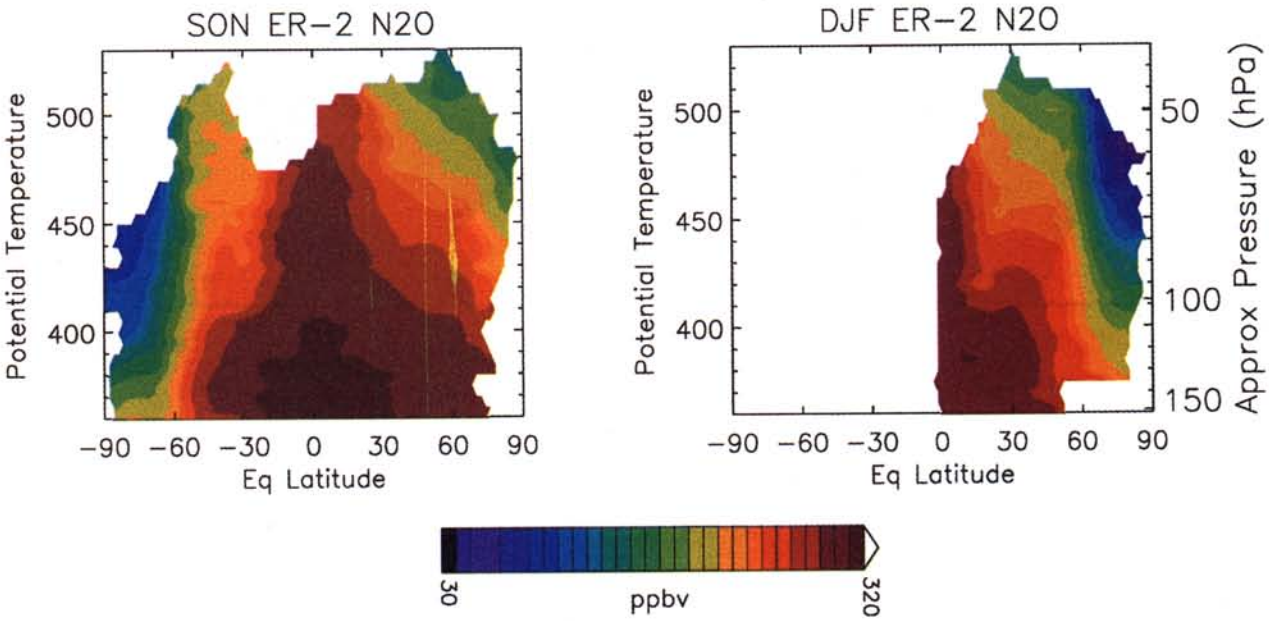

Plate 5. Contour plots of N_2O climatology in fall and winter derived from ATLAS measurements from the ER-2 during fall (left) and winter (right).

than that for the Northern Hemisphere. The vertical range of the observations is 360 - 530 K (about 150 - 50 hPa). The range 380 - 500 K has the most observations, so this comparison is limited to that range; observations are averaged in 20 K vertical bins.

Model mean profiles are calculated for the same seasons, latitude ranges, and potential temperatures as the ER-2 climatology. Because these comparisons are concerned with the transport above 380 K, the results are scaled so that the model mixing ratio is equal to the observed mixing ratio at 380 K (the tropical tropopause), the effective lower boundary for the stratosphere for this calculation. This small correction (a few percent) eliminates biases caused by model differences in tropospheric transport.

The fall (September-October-November) profiles from the models are compared with the climatological profiles in Figure 2. In the tropics the models values of N_2O are generally within the range of observations. The decrease with altitude is well represented. At northern

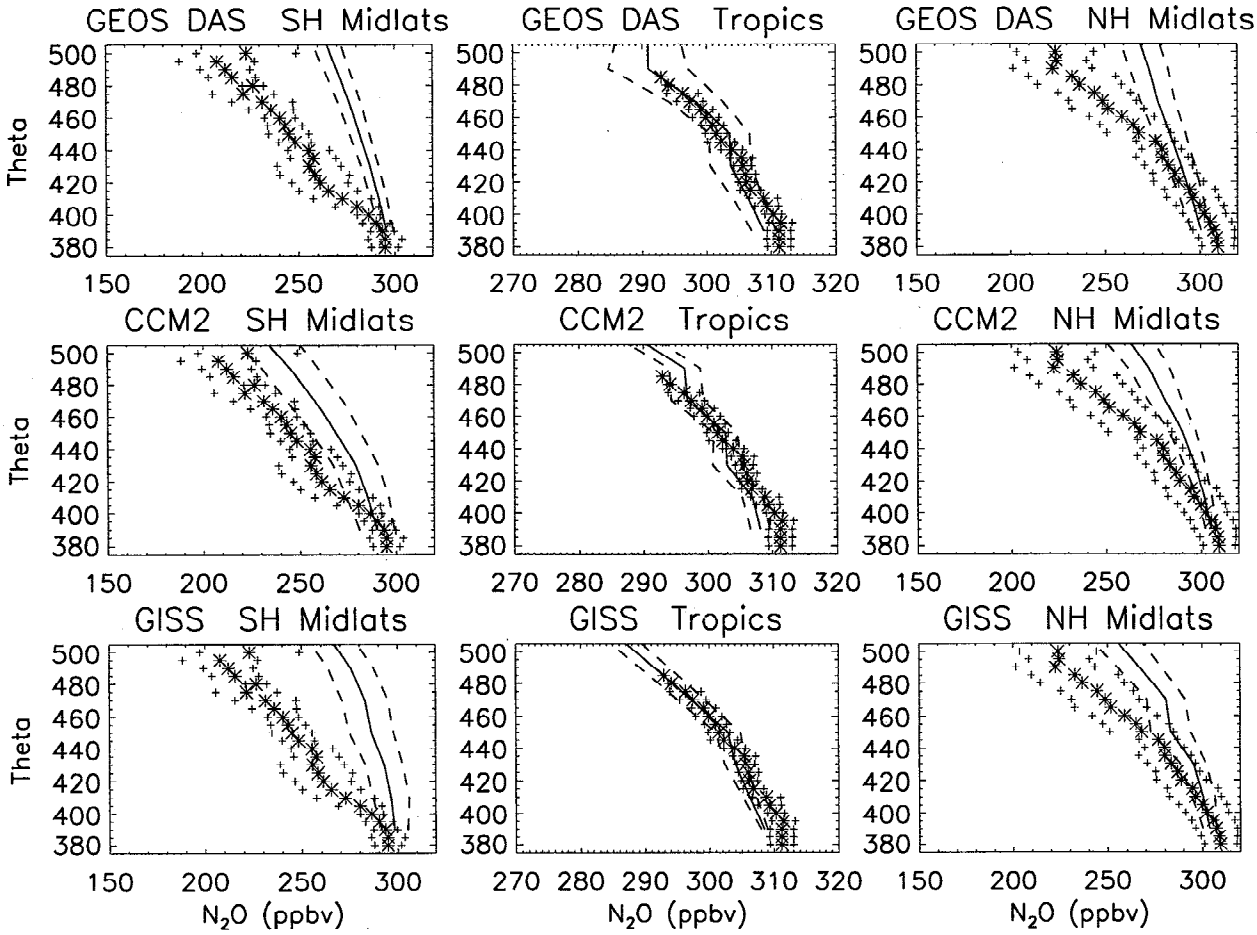

Figure 2. The fall (September-October-November) profiles from the models are compared with the climatological profiles. Each row corresponds to a different model (top: GEOS DAS; middle: CCM2; bottom: GISS) and each column to a different latitude range (left: SH middle; middle: tropics; right: NH middle).

middle latitudes the model profiles agree well with observed profiles in the lowest part of the stratosphere but diverge with increased potential temperature. At the highest potential temperature included in this comparison the model means are 270, 266, and 258 ppbv for GEOS DAS, CCM2, and GISS. All are significantly greater than the observed 224 ppbv. Results are similar at southern middle latitudes. Model means at 480 K are 270, 250, and 270 ppbv for GEOS DAS, CCM2, and GISS, compared with observed N_2O of 220 ppbv.

For each season the N_2O at each of the six potential temperatures in the climatological profile is compared with model N_2O. If the difference between the model and the climatology is within the standard deviation of the observations, the maximum score of 1 is given. A score of 0.5 is given if the difference is within the sum of the standard deviations of the model and the observations. No points are given for greater differences. There are 22 points of comparison (six potential temperatures x (3 seasons x 3 latitude ranges + 1 season x 2 latitude ranges)). The total score is scaled by the maximum possible score, so the final value of this test is 1. For this test, CCM2 receives a score of 0.68, GEOS DAS receives a score of 0.63, and GISS receives a score of 0.43. As for Test 2a, although model N_2O fields exhibit many characeristics of the ER-2 climatology, the differences are important. The models uniformly fail to reproduce the seasonal changes in N_2O between 440 and 500 K. Furthermore, the modeled N_2O is always higher than observations for this potential temperature range at middle latitudes. The discrepancies suggest that the horizontal mixing between the tropics and the middle latitudes is too vigorous relative to the modeled descent at middle latitudes and that the seasonal changes in the balance between the residual mean circulation and the horizontal mixing are poorly represented. N_2O and total reactive nitrogen NO_y exhibit a tight, linear anticorrelation in both observations and most models (not evaluated here). Therefore we anticipate that the systematic high bias in N_2O will be accompanied by a systematic low bias in NO_y.

4.5. Test 3 Tropical Midlatitude Distinctness of N_2O

Although the difference between the mean of all N_2O observations on a pressure level in the latitude band

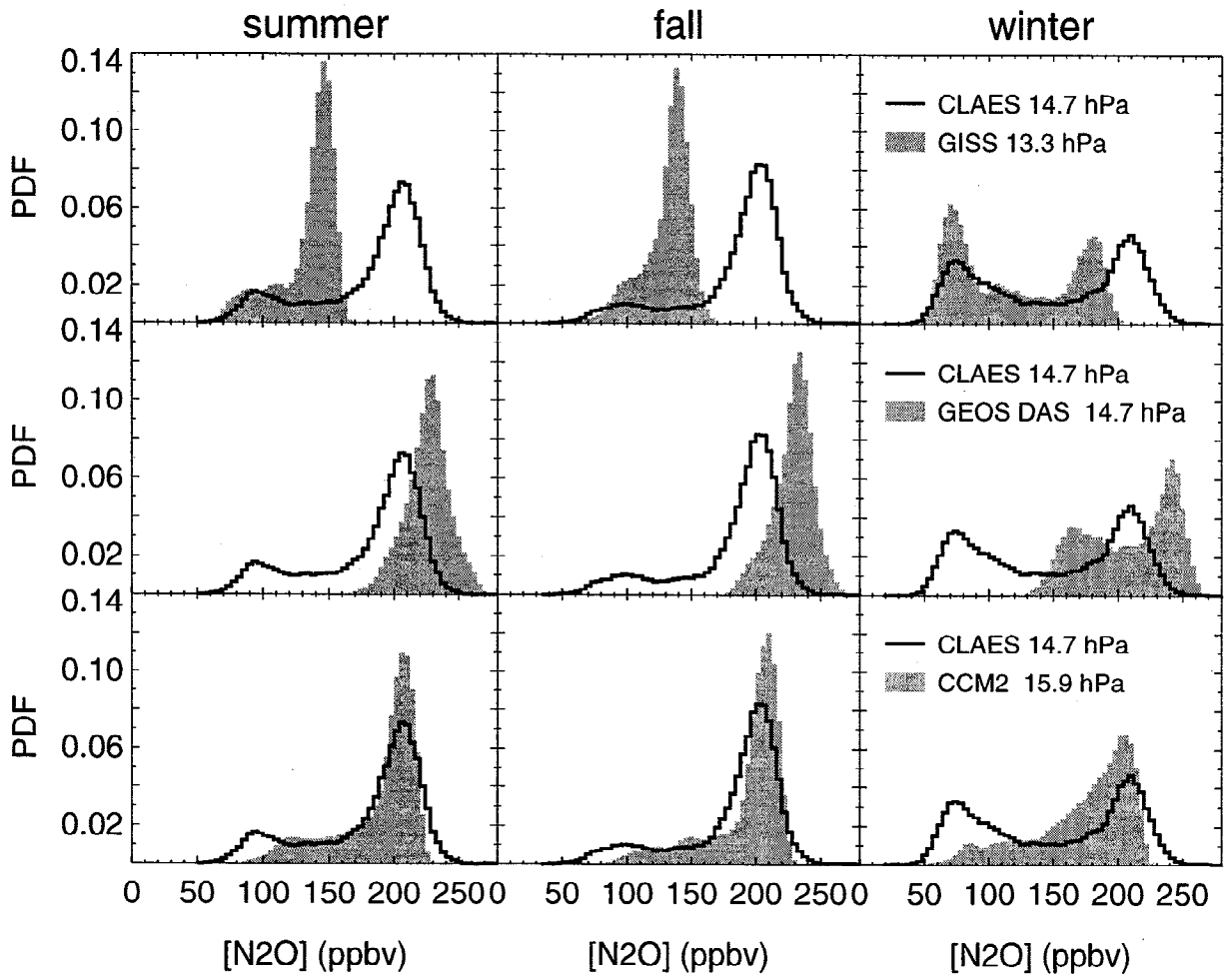

Figure 3. Seasonal equal-area histograms of CLAES N_2O at vertical level 14.7 hPa for summer (JJA 1992), fall (SON 1992), and winter (JF 1993). The shaded histograms are the model results, at vertical levels closest to the observations as indicated on the figure. Each row corresponds to a different model (top: GISS; middle: GEOS DAS; bottom: CCM2), and each column to a different season (left: summer; middle: fall; right: winter).

40°N to 50°N and that in the band 5°S to 5°N (section 2.1) provides a gross measure of the balance between transport by the mean winds and horizontal mixing, it provides little information about the sharpness and location of this tracer gradient. The meteorological processes responsible for mixing in and near the subtropics exhibit spatial and temporal variability, suggested by the differences between the fall and the winter distributions shown in Plate 2 (for example, note the change in the latitude of the maximum vertical gradient near 10 hPa). The sharpness of the horizontal gradient, as seen in Plate 2 (right) in the northern winter hemisphere, is produced by the combined influence of the residual circulation and the horizontal mixing driven by planetary wave transport. It contrasts with the more constant horizontal gradient seen in the northern fall hemisphere Plate 2 (left). Transport from the extratropics to the tropics has been the subject of recent analyses of satellite and aircraft observations (see references under Test 4). Somewhat less attention has been paid to transport in the opposite direction, i.e., tropics to extratropics [*Waugh*, 1996; *Boering et al.*, 1994]. There is outflow from the tropics at all levels associated with upward transport, and the tracer distribution at middle latitudes depends upon a balance between the mean horizontal and vertical transport as well as mixing across the subtropics.

One aspect of the large-scale structure in an observed or modeled N_2O field can be characterized by considering the shape of a probability distribution function (PDF), i.e., a histogram of the tracer field. The observations are weighted to account for nonuniform spatial sampling, and histograms for the model fields are computed by binning model values on a pressure surface which have been interpolated to an equal area grid (L. C. Sparling, Statistical perspectives on stratospheric transport, submitted to *Rev. Geophys.*, 1999]. PDFs were made for CLAES N_2O observations, using all mea-

surements between 10°S and 45°N at three UARS standard levels (31.6, 14.7, and 6.8 hPa) for summer (June, July, and August 1992), fall (September, October, and November 1992), and winter (January and February 1993). The distributions calculated for these time periods are stable; that is, they do not change appreciably if several days observations are omitted. December 1992 is not used in the compilation of winter statistics because there are large transients during this month, and the above condition is not satisfied. The northern spring distribution is strongly influenced by the breakup of the polar vortex, which varies greatly from year to year, and is omitted from this analysis. Also excluded is the summer 1992 distribution at 31.6 hPa, as the CLAES observations are uncertain due to the Mount Pinatubo aerosols. PDFs were made for each of the models, using the same latitude domain and the model pressure level closest to the UARS level. GEOS DAS pressures are the same as UARS pressures, GISS pressures are 23.7, 13.3, and 7.5 hPa, and CCM2 pressures are 31, 15.9, and 7.8 hPa. A total of eight distributions are compared. Results for 14.7 hPa are shown in Figure 3. The CLAES distributions are bimodal, that is, there are distinct tropical and midlatitude peaks in the mixing ratio distribution for each of the three seasons shown at 14.7 hPa, and for each season considered at the other two pressure levels as well. For each case, the model is given a score of 1 if the distribution reveals two distinct air masses (that is, if the distribution is bimodal) and zero otherwise. However, as Figure 3 shows, in the models the midlatitude mixing ratios often appear as a tail appended to the tropical peak, rather than the separate peaks seen by CLAES. Note that this test is based only on the structure of the distribution, and is a minimum requirement. The relative heights of the peaks, the magnitude of their separation, their seasonal behavior, and the altitude dependence of the distributions are graded in Test 2. If all of the models produced separate peaks, the scoring would be refined to reflect other areas of agreement or disagreement, for example, the relative population of the two distributions could be considered, or statistical tests to assess the agreement between the distributions could be applied. We anticipate refinement of this test as model fields are improved.

The combined score is normalized to a maximum of 1. The CCM2 N_2O distributions show two peaks for all eight cases, giving CCM2 a score of 1. GISS shows two peaks in five cases and receives a score of 0.62, and GEOS DAS shows distinct air masses in only three cases, receiving a score of 0.38.

4.6. Test 4 Propagation of Annual Cycle

One focus of model evaluation is transport in the lower tropical stratosphere. The rates of tropical entrainment of extratropical air and upwelling within the tropics determine the rate at which midlatitude lower stratosphere HSCT effluent can reach the ozone layer. The vertical propagation of an annual cycle in tracer mixing ratio, forced by variations at the tropopause, stringently tests model transport in this region. Examples of such tracers are CO_2 [*Boering et al.*, 1994, 1996] and $\hat{H}$=H_2O+2CH_4 [*Mote et al.*, 1996, 1998; *Randel et al.*, 1998]. The phase speed is not necessarily the same as the mean upwelling rate in the tropics [*Hall and Waugh*, 1997a; *Mote et al*, 1996], but in this case, the two speeds are approximately the same. The attenuation of the signal amplitude is related to tropical entrainment of extratropical air and vertically diffusive processes within the tropics [*Avallone and Prather*, 1996; *Minschwaner et al.*, 1996; *Volk et al.*, 1996; *Hall and Waugh*, 1997b; *Mote et al*, 1998]. Here properties of the modeled cycles are compared with estimates from in situ CO_2 and $\hat{H}$ observations taken during the Observations of the Middle Stratosphere (OMS) balloon campaign, (K. A. Boering et al., Timescales for stratospheric transport inferred from in situ observations of CO_2 from aircraft and balloons, manuscript in preparation, 1998), and HALOE $\hat{H}$, as analyzed by *Mote et al.* [1998] and *Randel et al.* [1998]. Data used here were provided by P. Mote and K. Boering. A similar comparison, employing a large number of models in addition to GMI, is made by *Hall et al.* [1999].

Figure 4 shows tropical profiles for the GMI models of the cycle amplitude profiles (natural log scale) and the phase lag time. (These profiles were derived from the model simulations of the age spectrum, as discussed by *Hall et al.* [1999].) Also plotted are the estimates from HALOE $\hat{H}$ data (crosses) and from in situ CO_2 (stars) and $\hat{H}$ data (diamonds). There is good agreement between in situ and satellite inferences of phase, and both show that the models propagate the annual cycle signal too rapidly; that is, the modeled phase lags increase too slowly with height.

For the amplitude the in situ inferences at the two highest altitudes are considerably less than the corresponding HALOE values. On the one hand, HALOE is known to underestimate the annual cycle in H_2O at the tropopause [e.g., *Mote et al.*, 1996] implying a possible overestimate of the fractional amplitude aloft. On the other hand, these top two in situ values are derived from the fewest number of balloon flights and may not be representative. In any case, the GMI model amplitudes bracket the observations. However, when proper account is taken for bias due to too rapid phase propagation (see below), all the models are seen to overattenuate the annual cycle.

Figure 4 shows that to a first approximation, models and measurements display simple exponential decay of amplitude (a straight line fit to the log amplitude) and uniform phase speed (a straight line fit to the phase lag). Thus the cycle propagation grade has two components, measured by single numbers: first, the average phase speed c from 16 km to 24 km; and second, the amplitude attenuation factor R. R is defined as H_a/λ, where H_a is the scale height of the exponential ampli-

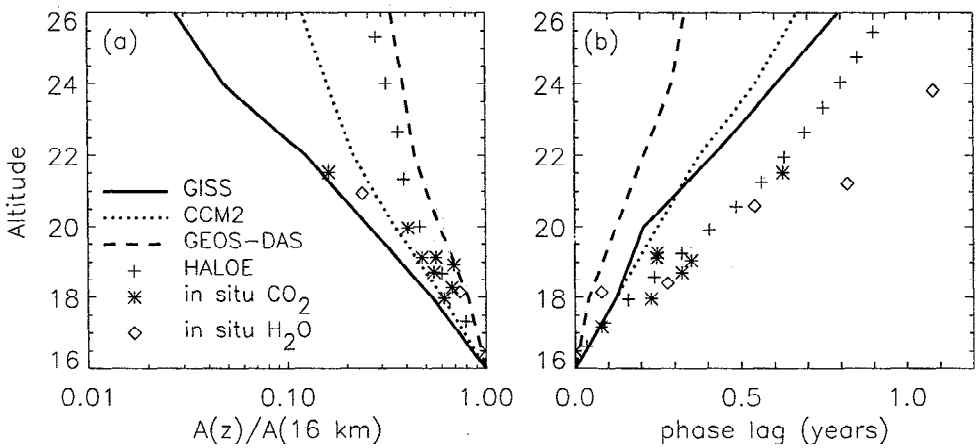

Figure 4. Vertical profiles in the tropics: (a) natural logarithm of the amplitude of the annual cycle in tracer mixing ratio propagating upward from the tropical tropopause; and (b) the phase lag time of the cycle. The amplitude is normalized to unity and the phase lag defined as zero at 16 km. Model profiles are (solid line) GISS, (short dashed) CCM2, and (long dashed) GEOS DAS. The symbols represent inferences from observations: (crosses) deductions from 5 years of HALOE-observed H_2O+2CH_4 data of *Mote et al.* [1998]; (stars) in situ CO_2 measurements from six aircraft and two balloon deployments from 1994 to 1997; and (diamonds) in situ H_2O+2CH_4 from the same two 1997 balloon deployments.

tude fit, and λ is the wavelength of the cycle (the phase speed times 1 year). R isolates the effectiveness of processes that attenuate the cycle better than H_a. As a measure of attenuation per wavelength, R summarizes the attenuation that occurs in each model over the same amount of time, 1 year. On a per wavelength basis, all the GMI models overattenuate the annual cycle.

Each model grade has three possible values: 1 if the model value for the attenuation factor R and the average phase speed c fall within the ranges for these values derived from the observations; 0.5 if the value is within 50% of either the upper or the lower bound of the observational range; and zero if the value is 50% greater than the upper bound or 50% less than the lower bound. Table 3 summarizes the values derived from the observations, the values derived for each model, the individual grades for phase and amplitude, and a mean of the two grades.

4.7. Test 5 Separation of the Upper Troposphere and Lower Stratosphere

This test concerns transport in the lowermost stratosphere, and the possibility of substantial vertical mixing between the lowermost stratosphere and the upper troposphere. *Nakazawa et al.* [1991], using a 2 year data set collected from commercial aircraft, report a CO_2 seasonal cycle in the upper troposphere (UT) near 60°N with a May maximum. At about the same latitude in the lower stratosphere (LS), 10-12 km, the observed amplitude of the CO_2 annual cycle was smaller, with a September maximum. These observations indicate a strong barrier to upward motion at the high-latitude tropopause, which is the basis for this test. The model tropopause is identified using model temperature fields. The tropopause is determined by the temperature minimum and/or a decrease in lapse rate to less than 2 K/km. For GEOS DAS and CCM2, which both have vertical resolution of about 1 km near the tropopause, the two model levels nearest the tropopause (identified using the above criteria) are considered the tropopause levels. For GISS, which has vertical resolution of about 3 km near the model tropopause, only one level is used. The model level just above the tropopause is considered "purely stratospheric," while the level just below is considered "purely tropospheric." The zonal mean of the CO_2 time series for each model at 60°N is evaluated. The observations show a 4 month difference between the maximum observed in the UT and that observed in the LS. A model is given a score of 1 if this difference is at least 2 months. While a model lag of 2 months is obviously significantly different from the observed lag, such a time interval is long enough to show that the seasonal cycle in the model lower stratosphere arrived via the tropical tropopause and not by vertical transport directly from the middle-latitude troposphere because the tropical to midlatitude lower stratospheric lag is less than 2 months [*Boering et al.*, 1994]. Thus satisfying these criteria is sufficient to claim that the pathway for model transport is realistic. All three models score 1 on this test when no convective transport is implemented. However, convective transport as implemented with the GEOS DAS winds and the CCM2 winds causes both models to fail this test, while the GISS winds, including convection, pass this test [*Hall and Prather*, 1993]. In the GISS GCM, convection is used to drive the general circulation, which provides self-consistency between the convection and the mean circulation. Unlike the HSCT assessment, the SASS evaluation will require a realistic

representation of convection, and this test will provide a strong constraint to the models.

4.8. Test 6 Horizontal Propagation of the CO_2 Signal

This analysis examines horizontal transport in the lower stratosphere but above the level of the tropical tropopause along quasi-potential temperature mixing surfaces. There are two aspects to this test. The first considers the amplitude of the annual cycle in CO_2, and the second considers the phase. As discussed by *Boering et al.* [1996] and *Strahan et al.* [1998], the extratropical CO_2 seasonal cycle between 380 and 440 K appears to be transported there from the tropics. Above ~440 K, no clear CO_2 seasonal cycle is observed in the Northern Hemisphere midlatitudes. To identify the cycle originating in the CO_2 at the tropical tropopause unambiguously, it is necessary to eliminate the component of the cycle due to the residual circulation (i.e., temperature changes in the lower stratosphere that cause seasonal variations in the relationship between CO_2 and potential temperature). This is accomplished by evaluating the CO_2 seasonal cycle on N_2O surfaces rather than potential temperature surfaces. Here the tropics are defined as 10°S to 10°N and the northern midlatitudes as 35°N to 55°N.

The amplitude test has two components. First, the model midlatitude seasonal cycle amplitude at 460 K must be less than 20% of the tropical seasonal cycle amplitude just above the tropopause (380 K). Second, the model seasonal cycle at 420 K in the midlatitudes must be at least 20% of the tropical seasonal cycle at 380 K. The models receive a score of 0.5 for each criterion which is satisfied. The GEOS DAS winds receive a score of 0.5; the CCM2 winds receive a score of 1. The GISS winds do not have sufficient native vertical resolution to make a meaningful application of this CO_2 amplitude test or the CO_2 phase test discussed in the following paragraph when run in the GMI CTM using the *Lin and Rood* [1996] numerical transport scheme. Applications using a transport scheme which conserves second-order moments resolves such vertical structures [*Prather,* 1986].

The CO_2 phase test, which is focused on the lower stratosphere at potential temperature greater than or equal to that at the tropical tropopause, considers the phase of the CO_2 seasonal cycle in the midlatitudes relative to its phase in the tropics. *Boering et al.* [1996] and *Strahan et al.* [1998] find that the observed midlatitude seasonal maximum between approximately 380 K and 460 K appears about 2 weeks after it appears in the tropics. This demonstrates that the primary pathway for transport from the midlatitude troposphere to the midlatitude lower stratosphere is through the tropics, rather than directly through the midlatitude tropopause. A model with a midlatitude seasonal maximum that arrives before the tropical CO_2 maximum implies an unrealistic pathway of transport in the model. For example, if the cycle maximum in the midlatitude LS arrived a few weeks after the maximum in the underlying troposphere, but before the arrival of the maximum in the tropics at the same height, this would imply unrealistic model transport directly up through the midlatitude tropopause. This test again uses N_2O as the vertical coordinate for CO_2. The N_2O bins (i.e., altitude ranges) are chosen to create three ranges between 380 and 460 K in the tropics. Three points are possible, one for each N_2O (altitude) range examined. The maximum score of 1 for each part of the test requires that the midlatitude seasonal maximum arrives 0.5-2 months after the tropical seasonal maximum. The score is 0.5 if the maxima appear within the half-month temporal resolution of the analysis and zero if the midlatitude maximum precedes the tropical maximum. The final score is the sum of the scores on each part divided by 3, so the maximum is 1. Both CCM2 and GEOS DAS receive a score of 1 on this part of the test, which indicates that horizontal transport is the dominant transport pathway. The GMI GISS results could not be diagnosed for this test. Results for both the CO_2 amplitude and the phase tests are summarized in Table 4.

5. Discussion and Conclusions

The overall scores for the three models are given in Table 5. The CCM2 meteorological fields equaled or outscored the GISS fields and equaled or outscored GEOS DAS fields on all tests except Test 1 (temperature) and were selected for the GMI assessment of the stratospheric HSCTs. Clearly, the process of grading and selection has just begun, and further comparisons will be part of a continuing model evaluation. For example, tracer tests are being extended to the full chemistry integrations (J. M. Rodriguez et al., The global modeling initiative assessment model: Testing of model results against observations of chemically-active species, manuscript in preparation, 1999). This compares the accuracy of these meteorological fields using the same chemical model with observed climatologies of column ozone and ozone profiles, with HALOE observations of NO and NO_2, and with an NO_y climatology for the lower stratosphere based on ER-2 data.

Although CCM2 is currently the best choice based on these tests, the tests also identify specific limitations to an assessment using CCM2 (as well as the GEOS DAS and GISS) meteorological fields. For example, because Northern Hemisphere winter temperatures are warmer than observed, the simulations with CCM2 fields will not predict polar stratospheric clouds, nor will they simulate any increase in PSC occurrence due to the build-up of H_2O and HNO_3 from the HSCTs.

A more important discrepancy identified here is the overestimate of N_2O in the lower stratosphere by all three models. Given the ready ability of most CTMs to match the observed N_2O-NO_y relationship in the lower stratosphere, it suggests that the models will signifi-

cantly underestimate absolute abundance of NO_y near 20 km, the peak expected HSCT perturbation to NO_y. The modeled ozone response to this increase in NO_y is itself dependent on the base level of NO_y; the ozone change per ppb increase in NO_y is less negative for lower background levels of NO_y [*Wennberg et al.*, 1994] as expected for these meteorological fields. A third important discrepancy is the inability of the models to represent all aspects of the separation of the tropics and midlatitudes, as shown by the PDFs in Test 3.

Nevertheless, the identification of these drawbacks and assignment of grades points out the benefit of this approach in evaluating and selecting models for an assessment. A model that scores as well as GEOS DAS on Test 1 would be expected to provide a more realistic representation of PSC occurrence and perturbation by HSCTs given an appropriate model for PSC formation and evaporation, [e.g., *Considine et al.*, 1999]. However, it is misleading with such a single test to assume that some of the more complex interactions of chemistry and transport in the polar stratosphere (e.g., the isolation of the PSC-processed polar vortex) would then be accurately simulated by the model. Similarly, a model with a higher score on Test 2 is likely to have a reasonable NO_y background in the lower stratosphere, but this does not necessarily mean that the buildup of HSCT NO_y would be more accurate. Provided we pick a set of grading criteria that give a balanced test of model performance in simulating atmospheric observations, the model with the highest score can be defended as the best choice for the current assessment.

Lastly, the establishment of formal criteria provides an objective baseline for evaluating new or improved versions of these meteorological fields in terms of their simulation of chemical tracers. As the general circulation models or data assimilation systems are developed and provide more realistic representations of stratospheric meteorology, these grades are likely to improve. However, neither GCMs nor data assimilation systems can simply be tuned to provide a better fit to these observations, thereby improving their grades. Simultaneous improvement in model performance on all these tests is not obvious (or likely). For example, developing a better parameterization for gravity-wave drag will affect both the residual circulation and the temperatures, and while the mean winds and temperature may better match observations, the aspects of tracer transport as tested here may not improve. A better representation of stratospheric meteorology, necessary to develop a credible assessment model for HSCTs, cannot be achieved by fitting transport parameters to a set of mean measurements. It is vital to establish grading criteria, such as used in this exercise, that represent the range of atmospheric phenomena controlling atmospheric chemistry.

Acknowledgments. We thank Kristie Boering and Phil Mote for providing additional analysis of their data sets, Mark Schoeberl for providing artistic talent in preparing Plate 1, and the rest of the GMI team for providing critical comments and suggestions throughout this project. This work was supported by the High Speed Research Program of NASA's Atmospheric Effects of Aviation Project.

References

Albritton, D. L., et al., The atmospheric effects of stratospheric aircraft: Interim assessment report of the NASA high-speed research program, *NASA Ref. Publ. 1333*, 1993.

Andrews, D. G., J. R. Holton, and C. B. Leovy, *Middle Atmosphere Dynamics*, 489 pp., Academic, San Diego, Calif., 1987.

Avallone, L. M., and M. J. Prather, Photochemical evolution of ozone in the lower tropical stratosphere, *J. Geophys. Res., 101*, 1457-1461, 1996.

Avallone, L. M., and M. J. Prather, Tracer-tracer correlations: Three-dimensional model simulations and comparisons to observations, *J. Geophys. Res., 102*, 19,233-29,246, 1997.

Boering, K. A., B. C. Daube, Jr., S. C. Wofsy, M. Loewenstein, J. R. Podolske, and E. R. Keim, Tracer-tracer relationships and lower stratospheric dynamics: CO_2 and N_2O correlations during SPADE, *Geophys. Res. Lett., 21*, 2567-2570, 1994.

Boering, K. A., S. C. Wofsy, B. C. Daube, J. R. Schneider, M. Loewenstein, J. R. Podolske, and T. J. Conway, Stratospheric mean ages and transport rates from observations of CO_2 and N_2O, *Science, 274*, 1340-1343, 1996.

Boville, B. A., Middle atmosphere version of CCM2 (MACCM2): Annual cycle and interannual variability, *J. Geophys. Res., 100*, 9017-9039, 1995.

Brasseur, G. P., X. X. Tie, P. J. Rasch, and F. Lefevre, A three-dimensional simulation of the Antarctic ozone hole: Impact of anthropogenic chlorine on the lower stratosphere and upper troposphere, *J. Geophys. Res., 100*, 8909-8930, 1997.

Considine, D. B., A. R. Douglass, P. S. Connell, D. E. Kinnison, and D. A. Rotman, A polar stratospheric cloud parameterization for the three-dimensional model of the global modeling initiative and its response to stratospheric aircraft emissions *J. Geophys. Res.*, in press, 1999.

Coy, L., and R. Swinbank, Characteristics of stratospheric winds and temperatures produced by data assimilation, *J. Geophys. Res., 102*, 25,763-25,781, 1997.

Coy, L., R. B. Rood, and P. A. Newman, A comparison of winds from the STRATAN data assimilation system to balanced wind estimates, *J. Atmos. Sci., 51*, 2309–2315, 1994.

Douglass, A. R., M. A. Carroll, W. B. DeMore, J. R. Holton, I. S. A. Isaksen, H. S. Johnston, and M. K. W. Ko, The atmospheric effects of stratospheric aircraft: A current consensus, *NASA Ref. Publ. 1251*, 1991.

Douglass, A. R., R. B. Rood, C. J. Weaver, M. C. Cerniglia, and K. F. Brueske, Implications of thre-dimensional tracer studies for two-dimensional assessments of the impact of supersonic aircraft on stratospheric ozone, *J. Geophys. Res.,98*, 8949-8963, 1993.

Douglass, A. R., C. J. Weaver, R. B. Rood, and L. Coy, A three-dimensional simulation of the ozone annual cycle using winds from a data assimilation system, *J. Geophys. Res., 101*, 1463-1474, 1996.

Graedel, T. E., D. Cariolle, M. A. Geller, J. Kerrebrock, D. Lister, K. Mauersberger, S. Penkett, U. Schmidt, S. Schwartz, and S. Solomon, *Atmospheric Effects of Stratospheric Aircraft: An Evaluation of NASA's Interim Assessment*, 45 pp., Natl. Acad. Press, Washington, D. C., 1994.

Gray, L. J., and J. M. Russell III, Interannual variability of trace gases in the subtropical winter stratosphere, *J. Atmos. Sci., 56*, 977-993, 1999.

Hack, J. J., B. A. Boville, J. T. Kiehl, P. J. Rasch, and D. L. Williamson, Climate statistics from the National Center for Atmospheric Research community climate model CCM2, *J. Geophys. Res., 99*, 20,785-20,813, 1994.

Hall, T. M., and M. J. Prather, Simulations of the trend and annual cycle in stratospheric CO_2, *J. Geophys. Res., 98*, 10,573-10,581, 1993.

Hall, T. M., and M. J. Prather, Seasonal evolutions of N_2O, O_3, and CO_2: Three-dimensional simulations of stratospheric correlations, *J. Geophys. Res., 100*, 16,699-16,720, 1995.

Hall, T. M. and D. W. Waugh, Timescales for the stratospheric circulation derived from tracers, *J. Geophys. Res., 102*, 8991-9001, 1997a.

Hall, T. M. and D. W. Waugh, Tracer transport in the tropical stratosphere due to vertical diffusion and horizontal mixing, *Geophys. Res. Lett., 24*, 1383-1386, 1997b.

Hall, T. M, D. W. Waugh, K. A. Boering, and R. A. Plumb, Evaluation of transport in stratospheric models, *J. Geophys. Res., 104*, 18,815-18,839, 1999.

Hannegan, B., S. Olsen, M. Prather, X. Zhu, D. Rind, and J. Lerner, The dry stratosphere: A limit on cometary water influx, *Geophys. Res. Lett., 25*, 1649-1652, 1998.

Hansen, J., G. Russell, D. Rind, P. Stone, A. Lacis, S. Lebedeff, R. Ruedy, and L. Travis, Efficient 3-D global models for climate studies: Models I and II, *Mon. Weather Rev., 111*, 609-662, 1983.

Lin, S. J., and R. B. Rood, Multidimensional flux form semi-Lagrangian transport schemes, *Mon. Weather Rev., 124*, 2046-2070, 1996.

McIntyre, M. E., Atmospheric dynamics: Some fundamentals, with observational implications, in *The Use of EOS for Studies of Atmospheric Physics*, edited by J. C. Gille and G. Visconti, pp. 313-386, North-Holland, New York, 1992.

Minschwaner, K., A. E. Dessler, J. W. Elkins, C. M. Volk, D. W. Fahey, M. Loewenstein, J. R. Podolske, A. E. Roche, and K. R. Chan, Bulk properties of isentropic mixing into the tropics in the lower stratosphere, *J. Geophys. Res., 101*, 9433-9440, 1996.

Mote, P. W., K. H. Rosenlof, M. E. McIntyre, E. S. Carr., J. C. Gille, J. R. Holton, J. S. Kinnersley, H. C. Pumphrey, J. M. Russell III, and J. W. Waters, An atmospheric tape recorder: The imprint of tropical tropopause temperatures on stratospheric water vapor, *J. Geophys. Res., 101*, 3989-4006, 1996.

Mote, P. W., T. J. Dunkerton, M. E. McIntyre, E. A. Ray, P. H. Haynes, and J. M. Russell, Vertical velocity, vertical diffusion, and dilution by midlatitude air in the tropical lower stratosphere, *J. Geophys. Res., 103*, 8651-8666, 1998.

Nakazawa,T., K. Miyashita, S. Aaki, and M. Tanaka, Temporal and spatial variations in upper tropospheric and lower stratospheric carbon dioxide, *Tellus, Ser. B, 43*, 106-117, 1991.

O'Sullivan, D., and T. J Dunkerton, The influence of the quasi-biennial oscillation on global constituent distributions, *J. Geophys. Res., 102*, 21,731-21,743, 1997.

Park, J. H., M. K .W. Ko, C. H. Jackman, and R. A. Plumb, Models and Measurements II, *NASA Ref. Publ.*, in press, 1999.

Podolske, J. R., and M. Loewenstein, Airborne tunable diode laser spectrometer for tracer gas measurement in the lower stratosphere, *Appl. Opt., 32*, 5324-5333, 1993.

Prather, M. J., Numerical advection by conservation of second-order moments, *J. Geophys. Res., 91*, 6671-6681, 1986.

Prather, M. J., and E. E. Remsberg (Eds.), The atmospheric effects of stratospheric aircraft: Report of the 1992 models and measurements workshop, *NASA Ref. Publ. 1292*, 1993.

Prather, M. J., H. L. Wesoky, R. C. Miake-Lye, A. R. Douglass, R. P. Turco, D. J. Wuebbles, M. K. W. Ko, and A. L. Schmeltekopf, The Atmospheric effects of stratospheric aircraft: A first program report, *NASA Ref. Publ. 1272*, 1992.

Randel, W. J., B. A. Boville, J. C. Gille, P. L. Bailey, S. T. Massie, J. B. Kumer, J. L. Mergenthaler, and A. E. Roche, Simulation of stratospheric N_2O in the NCAR CCM2: Comparison with CLAES data and global budget analyses, *J. Atmos. Sci., 51*, 2834-2845, 1994.

Randel, W. J., F. Wu, J. M. Russell, A. Roche, and J. W. Waters, Seasonal cycles and QBO variations in stratospheric CH_4 and H_2O observed in UARS HALOE data, *J. Atmos. Sci., 55*, 163-184, 1998.

Rasch, P. J., X. Tie, B. A. Boville, and D. L. Williamxon, A three-dimensional transport model for the middle atmosphere, *J. Geophys. Res., 99*, 999-1017, 1994.

Rasch, P. J., B. A. Boville, and G. P. Brasseur, A three-dimensional general circulation model with coupled chemistry for the middle atmosphere, *J. Geophys. Res., 100*, 9041-9071, 1995.

Rind, D., and J. Lerner, The use of on-line tracers as a diagnostic tool in GCM model development, *J. Geophys. Res., 101*, 12,667-12,683, 1996.

Rind, D., D. Shindell, P. Lonergan, and N. K. Balachandran, Climate change and the middle atmosphere, III, The doubled CO_2 climate revisited, *J. Clim., 11*, 876-894, 1998.

Roche, A. E., J. B. Kumer, J. L. Mergenthaler, G. A. Ely, W. G. Uplinger, J. F. Potter, T. C. James, and L. W. Sterritt, The cryogenic limb array etalon spectrometer (CLAES) on UARS: Experiment description and performance, *J. Geophys. Res., 98*, 10,763-10,775, 1993.

Roche, A. E., et al., Validation of CH_4 and N_2O measurements by the cryogenic limb array etalon spectrometer instrument on the Upper Atmosphere Research Satellite, *J. Geophys. Res., 101*, 9679-9710, 1996.

Russell, J. M., III, L. L. Gordley, J. H. Park, S. R. Drayson, A. F. Tuck, J. E. Harries, R. J. Cicerone, P. J. Crutzen, and J. E. Frederick, The Halogen Occultation Experiment, *J. Geophys. Res., 98*, 10,777-10,797, 1993.

Schubert, S. D., R. B. Rood, and J. Pfaendtner, An assimilated data set for Earth science application, *Bull. Am. Meteorol. Soc., 74*, 2331-2342, 1993.

Stolarski, R. S., et al., 1995 Scientific Assessment of the atmospheric effects of stratospheric aircraft, *NASA Ref. Publ. 1381*, 1995.

Strahan, S. E., A. R. Douglass, J. E. Nielsen, and K. A. Boering, The CO_2 seasonal cycle as a tracer of transport, *J. Geophys. Res., 103*, 13,729-13,741, 1998.

Strahan, S. E., M. Loewenstein and J. R. Podolske, Climatology and small-scale structure of lower stratospheric N_2O based on in situ observations, *J. Geophys. Res., 104*, 2195-2208, 1999.

Volk, C. M., et al., Quantifying transport between the tropical and mid-latitude lower stratosphere, *Science, 272*, 1763-1768, 1996.

Waugh, D. W., Seasonal variation of isentropic transport out of the tropical stratosphere, *J. Geophys. Res., 101*, 4007-4023, 1996.

Weaver, C. J., A. R. Douglass, and R. B. Rood, Thermodynamic balance of three-dimensional stratospheric winds derived from a data assimilation procedure, *J. Atmos. Sci., 50*, 2987-2993, 1993.

Weaver, C. J., A. R. Douglass, and R. B. Rood, Tracer transport for realistic aircraft emission scenarios calculated using a three-dimensional model, *J. Geophys. Res., 100*, 5203-5214, 1995.

Weaver, C. J., A. R. Douglass, and D. B. Considine, A 5-year simulation of supersonic aircraft emission transport using a three-dimensional model, *J. Geophys. Res., 101*, 20,975-20,984, 1996.

Wennberg, P. O., et al., The removal of lower stratospheric O_3 by free radical catalysis: In situ measurements of OH, HO_2, NO, NO_2, ClO, and BrO, *Science, 166*, 398-404, 1994.

L. Coy, A. Douglass, L. Sparling, and S. Strahan, NASA Goddard Space Flight Center, Code 916, Greenbelt, MD 20771. (e-mail: coy@demeter.gsfc.nasa.gov, douglass@persephone.gsfc.nasa.gov, sparling@dynarama.gsfc.nasa.gov, sstrahan@dao.gsfc.nasa.gov)

T. M. Hall, NASA Goddard Institute for Space Studies, 2880 Broadway, New York, NY 10025. (e-mail: hall@giss.nasa.gov)

M. J. Prather, Earth System Science Department, University of California, Irvine, CA 92717. (e-mail: mprather@uci.edu)

P. J. Rasch, Climate and Global Dynamics Division, National Center for Atmospheric Research, Boulder, CO 80307. (e-mail: pjr@ncar.ucar.edu)

J. M. Rodriguez, Department of Marine and Atmospheric Chemistry, Rosenstiel School of Marine and Atmospheric Science, University of Miami, 4600 Rickenbacker Cuaseway, Miami FL 33149. (email: jrodriguez@rsmas.miami.edu)

(Received December 22, 1998; revised July 19, 1999; accepted July 28, 1999.)